CDI SSF Elements: Science Data Lifecycle Model, Computational Tools and Services

Data Management Training Clearinghouse

Principal Investigator: JC Nelson, USGS Upper Midwest Environmental Science Center, 2630 Fanta Reed Road La Crosse WI 54603, 608-781-6370, Fax 608-783-6066, jcnelson@usgs.gov
Additional PIs:
Nancy Hoebelheinrich, Knowledge Motifs LLC, San Mateo, CA 94401, 650-302-4493, Fax 650-745-3333, nhoebel@kmotifs.com
Tamar Norkin, USGS Core Science Analytics, Synthesis and Libraries, Denver Federal Center, Bldg. 810, MS 302, Lakewood, CO 80225, 303-202-4229, Fax 303-202-4229, tnorkin@usgs.gov,
Fiscal Contact: Marissa Borre, USGS Upper Midwest Environmental Science Center, 2630 Fanta Reed Road La Crosse WI 54603, p. 608-781-6303, Fax 608-783-6066, mborre@usgs.gov
Collaborators: Amber Budden (DataONE, aebudden@dataone.unm.edu), John Faundeen (USGS EROS, faundeen@usgs.gov), Sophie Hou (University of Illinois, Urbana-Champaign, hou@illinois.edu), Shelley Knuth (University of Colorado, Boulder, Shelley.Knuth@colorado.edu), Matt Mayernik (NCAR, mayernik@ucar.edu), Erin Robinson (Federation of Earth Science Information Partners, erinrobinson@esipfed.org), David Bassendine (Blue Dot Lab, david@bluedotlab.org).

Abstract: Many organizations, government agencies and academic institutions have developed excellent educational resources to inform and train scientists about best practices in data management. It has been our experience that educational resources are often difficult to locate, as they are spread out across numerous websites and scientific domains. In addition, resources are often difficult to evaluate with respect to their effectiveness and their relationship to a specific domain or framework. To address this need, a group of organizations that have experience creating, presenting and distributing data management training resources will collaborate to create a clearinghouse for data management training resources in the Earth sciences. The clearinghouse will provide a searchable and browseable inventory of training resources and will be hosted within the Federation for Earth Science Information Partners' (ESIP's) existing Drupal environment at esipfed.org. To assist with discoverability and scope definition, training resources will be associated with elements of the USGS Science Data Lifecycle Model. The team will develop a sustainable mechanism for adding and describing new resources through crowdsourcing. An advisory group of interested and committed organizations has been formed to guide and support the project; represented organizations include the Federation of Earth Science Information Partners, DataONE, USGS Core Science Analytics, Synthesis and Libraries, and Knowledge Motifs LLC. The advisory group will continue its work beyond the grant period to sustain the effort using resources outside the scope of this proposal.

Total funding amount requested: $47,834.70
Total in-kind funding: $16,386

Expected Products Generated: online clearinghouse of data management training resources, template and online help guide to facilitate the description of training resources, final report that contains feasibility study for next steps.

General Public Summary

Many organizations, government agencies and academic institutions have developed excellent educational resources to inform and train research scientists about best practices in data management. These educational resources are often difficult to locate, however, as they are spread out across a wide variety of websites and scientific domains. In addition, it's often difficult for scientists to determine which training resources would be most helpful for their specific needs. A group of organizations that have experience creating, presenting and distributing data management training resources will collaborate to create a clearinghouse for data management training in the Earth sciences. The clearinghouse will allow scientists to search and browse an inventory of training resources related to data management in the Earth sciences. The clearinghouse will also allow users to enter and describe new resources, so that the inventory can continue to grow and stay up-to-date. An advisory group of organizations has been formed to guide and support the project; represented organizations include the Federation of Earth Sciences, DataONE, USGS Core Science Analytics, Synthesis and Libraries, and Knowledge Motifs LLC.

Proposal Narrative

Scope

Many organizations, government agencies and academic institutions in the U.S. and abroad have developed excellent learning resources to inform and train scientists and their information partners about best practices in data management. Some of these learning resources are focused on data management within specific domains. It has been our experience, however, that these kinds of learning resources are often difficult to locate, as they are spread out across numerous websites and may be available from domains that research scientists of a given discipline may not think to look. For example, the data management training modules found on the USGS Data Management website provide instruction of a general nature but are not likely to be discovered by researchers outside of the USGS. In addition, learning resources are often difficult to evaluate and compare with respect to their effectiveness and their relationship to a specific domain-relevant data management framework. To address this need, a group of organizations with experience in creating, presenting and distributing data management learning resources will collaborate to establish a Data Management Training Clearinghouse ("the Clearinghouse") for learning resources on data management in the Earth Sciences.

This project will build on previous USGS Community for Data Integration (CDI) funded projects, including the Science Data Lifecycle Model (part of the CDI Science Support Framework) and the USGS Data Management website. Although both of these projects were focused within the USGS, the proposed project will have a broad scope and will involve collaboration with academic institutions and non-governmental organizations and communities. An advisory group of interested and committed organizations has been formed to guide and support the project; represented organizations include the Federation of Earth Sciences, DataONE, USGS Core Science Analytics, Synthesis and Libraries, and Knowledge Motifs LLC. Members of the advisory group have experience in creating and presenting training resources (for example, the DataONE education modules and the ESIP Data Management Short Course for Scientists). The advisory group will continue its work beyond the grant period to sustain the effort.

The Clearinghouse project team is operating on the assumption that scientific researchers usually prioritize their time more on pursuing their research than on managing the data that they acquire, process, and publish[1]. As a result, one of the main objectives of the effort is to make it much easier for Earth science researchers to quickly find learning resources on data management for their own use, for the use of their research team, for passing along to their students. Another objective is to help scientists assess the learning resources that they find, in order to ascertain the context in which they were created. For example, it would be useful for a potential learning resource user to know which organization created the learning resource, the type of resource (e.g., static presentation, video with audio file, online tutorial with exercises, etc.), the educational or data management framework to which the resource aligns, the target audience, the length of time needed to use the resource, and, possibly, ratings by previous users. Information about the potential for re-use could be quite helpful, e.g., by describing the licensing context that allows resources to be incorporated into other, project-specific training resources. By quickly

---

[1] Rüegg, J., Gries, C. Bond-Lamberty, B., Bowen, G.J., Felzer, B.S., McIntyre, N.E., Soranno, P.A., Vanderbilt, K.L., and Weathers, K.C., 2014: Completing the data life cycle: using information management in macrosystems ecology research. *Frontiers in Ecology and the Environment* **12**: 24–30. http://dx.doi.org/10.1890/120375.

reviewing these kinds of details about a learning resource, a researcher's time can be more efficiently spent going directly to the resources that best align with their specific needs.

The organizing group believes that it would be helpful to associate learning resources with elements of a data management framework, and intends to use the USGS Science Data Lifecycle (part of the CDI Science Support Framework) for this purpose. (See http://www.usgs.gov/cdi/cdi-ssf/cdi-ssf-components.pdf.) This will allow researchers to hone in on the most relevant resources, especially if they are interested in finding guidance on a specific data management task. This could also be helpful in highlighting the steps of the lifecycle and communicating the idea that data management should be incorporated into all stages of the research process, from planning to publishing. Understanding the relationship of one step of the Science Data Lifecycle to the others can give the researcher a better perspective on the iterative nature of the Science Data Lifecycle steps and, hopefully, enhance the motivation to engage in better data management.

As the Clearinghouse develops, there may be other educational frameworks or descriptions of data lifecycles that could be applied, such as the DataONE Data Life Cycle. The Clearinghouse will be set up in such a way that more metadata of this nature to include different types of lifecycles could be added as desired.

One of the key impacts of the Clearinghouse as we envision it is to simply help Earth Science researchers find appropriate learning resources on data management. Tagging the learning resources with search community approved Schema.org metadata elements will facilitate generic search engine discovery within Google, Microsoft, Yahoo, and Yandex. As an additional consequence of following the Schema.org metadata elements, the contextual / descriptive information about the learning resources can be available for harvesting and distributing by Semantic web and other web-based services by virtue of using the Learning Resource Metadata Initiative (LRMI) metadata elements. Thus, the inclusion of metadata elements designed specifically to feature educational aspects of the resources should greatly facilitate the identification of appropriate resources, and also allow learning resource users to take different approaches to learning, such as just-in-time learning, or sequentially through a data lifecycle process.

Another impact of the Clearinghouse would be upon the organizations that create data management learning resources, including the USGS CDI, DataONE, and the Federation of Earth Science Information Partners (ESIP). Making the learning resources available via a Clearinghouse would not only extend the reach of the organizations to their own constituents, but would also broaden and deepen the range of learning resources available to all constituents by pulling from communities of practice outside their own domains. Also, we should be able to better identify gaps in the topics covered by existing learning resources on data management, and inform the development of additional learning resources.

The latter impact, identifying the gaps in training topics and informing further development of training topics, will be accomplished by expanding upon a Data Management Training Resources Survey, a student fellow project conducted by ESIP Student Fellow, Sophie Hou, in Fall, 2015. (See http://commons.esipfed.org/node/8763). This survey compiled and compared a list of topics on data management from nine different types of U.S. and international organizations, including federal science agencies, research consortia, and university libraries.

The survey explored the extant "landscape" of the data management training resources with Earth science as the targeted science domain. The survey was proposed to achieve two key goals. First, the survey

results were intended to allow the survey's sponsor, the ESIP Data Stewardship Committee, to better understand the types of training resources and content delivery formats that are currently available to support data management training. Second, selected training resources were compared in order to identify commonalities and gaps in the training topics discussed.

The survey results will be helpful to further evaluate, develop, and expand the roadmap for creating the next generation of data management training resources. In addition, the survey results can be usefully applied to the planning, structuring, and production of the Clearinghouse project in several ways. First, the Clearinghouse organizing group can use the list of learning resources topics / titles as an initial inventory of learning resources that can be added to the Clearinghouse. Second, the comparison of the topics covered by the nine different organizations reveals a set of topics that could be considered "core knowledge" for the growing discipline of data management training. The core topical areas could be used to ensure that the Clearinghouse tracks, compares and includes learning resources on these topic both initially and over time. The fact that these topics were determined independently by differing sets of experts further solidifies the core knowledge topics. Third, the topics that were not covered as regularly by the existing training resources (but were appearing more frequently in the data management community's discussions) could be used as topics for which learning resources could be sought (or developed by collaborating organizations) as time allows in the first phase, and in the future phases of Clearinghouse development. The survey provides many topics that were considered and presented by the broader community to be relevant and crucial for performing data management tasks, but were neither "core" nor "trending". As a result, the capture of these training topics and resources could be shared and accessed by a wide ranging audience by virtue of inclusion in the initial inventory of learning resources in the Clearinghouse. Examples of these topics included data quality, cost model, and tool recommendations for data management.

Technical Approach

Summary and Key Features:
Our overall technical objective for the Clearinghouse is to develop a sustainable, crowd-sourced, web-based metadata repository for online learning resources focused on data management training. In future, we will also track face-to-face events and activities.

Visitors will be able to:
- discover resources by browsing steps in the USGS Science Data Lifecycle or running searches
- discover resources when external search engines pick up learning resource pages, which are well structured for SEO and (in future) will provide LRMI schema.org metadata
- filter or sort resource listings by metadata field values, e.g. media type or submitting organization
- suggest new resources to include, using a streamlined form with inline contextual help

Editors and Peer Reviewers will be able to:
- receive notifications when learning resource submissions arrive
- choose, read, and change the metadata entered as necessary, comparing revisions
- send the final drafts of metadata record to editors after peer review or an agreed time period
- publish reviewed learning resources, notifying Clearinghouse subscribers when they go live
- review and fix broken links, automatically checked every day and flagged in an admin dashboard

Architecture for rapid prototyping and long-term sustainability:

The Clearinghouse will be a new subsite within the ESIP family, accessible directly at its own subdomain or from the "Explore" menu – similar to the existing Testbed (http://testbed.esipfed.org) and Commons (http://commons.esipfed.org). The content of the Clearinghouse will be stored in ESIP's Rackspace cloud storage, via Pantheon (http://pantheon.io), a Drupal specific hosting platform.

We will inherit a rich Drupal architecture that will jump start development and allow us to focus on work that really adds value. From day one, we'll benefit from ESIP's existing:

- Pantheon hosting environment – optimized for Drupal with Varnish caching for scalability, Apache Solr for indexing and querying content, and New Relic for real-time monitoring
- Drupal architecture – reuse existing core Drupal, contributed modules, content types and search
- Bootstrap frontend framework – to streamline design, styling and device and browser testing

In future, we will leverage Drupal's contributed modules to pull learning resources directly from upstream repositories (sustainably widening our local catalog) and push ours back (boosting discovery in the wider community, e.g. to http://learningregistry.org using https://drupal.org/project/learning_registry)

Work needed and milestones:
If the Clearinghouse project is funded, we will work through the following series of feature-based milestones that establish workflows, schemas and help content; set up Drupal architecture; lay out responsive Bootstrap interfaces; then test across mobile, tablet and desktop devices.
1. Discovery – rapid discovery to establish overall patterns for the application
   a. Information Architecture – page types, content and sitemap
   b. User Roles and Stories – define roles and outline key workflows
   c. Wireframing – identify page elements and map them to responsive layout

2. Development – Drupal architecture and code to provide the following features:
   d. Set up ESIP microsite – set up ESIP's existing architecture to provide a new subdomain
   e. Learning Resource – standards-based descriptive metadata field schema, page layout with branding for contributing organizations (eg USGS, DataONE, and others)
   f. Search and Listings – main faceted search interface, providing a learning resource listing that's filterable by defined vocabularies (eg. topic or type) or open keyword search
   g. Homepage – adjust the main interface for introductions, tips and learning resource highlights
   h. Crowdsource Learning Resources – allow authenticated visitors to submit suggestions, providing inline help for metadata fields, and a context-sensitive glossary for terms
   i. Moderate Crowdsourced Learning Resources – develop criteria and timelines for peer review and editor metadata evaluations; set up workflow using workflow and rules modules
   j. Automated link checking – set up Link Checker module
   k. Deployment – final testing, then publish our Phase 1 application to ESIP's live environment
   l. Content entry – populate the Clearinghouse just after release, drawing from the ESIP Data Stewardship Committee's Survey (inc. USGS Data Management & DataONE modules)

Project Experience and Collaboration

PI: JC Nelson:

PI: Nancy J. Hoebelheinrich:
Nancy has extensive experience working and collaborating in many guises, i.e, as a project manager, an editor, a co-creator of educational resources, and as a liaison between information users and software

developers. Her involvement in these kinds of activities has ranged from volunteer to paid staff, and for different kinds of organizations ranging from non-profit, academic, and for profit organizations. She has been involved with metadata standards development as well as software and web services development over her 25 or so years working in the information industry. Most recently, she has been Chair of the ESIP Data Management Training Working Group, Co-editor of the ESIP Data Management Short Course for Scientists and author of several of those modules. Some of the products of her participation in recent collaborative projects include the development of metadata de-jure and de-facto metadata standards, XML and RDF ontologies leading to web based services, and metadata inputting forms using XForms and Drupal.

PI: Tamar Norkin:
Tamar works in the Science Data Management branch of the U.S. Geological Survey's Core Science Analytics, Synthesis and Libraries program, which creates and manages a set of tools and online guidance material to assist USGS scientists with data management. As a member of this team, Tamar has experience assisting researchers and data managers with the USGS Science Data Lifecycle steps and helping them meet new USGS open data requirements. She is the point of contact for USGS scientists who release their data through ScienceBase, an online data platform. Tamar is a member of CDI's Data Management Working Group and helps coordinate a monthly webinar on data management topics.

Contributors: John Faundeen, Amber Budden, Sophie Hou, Matt Mayernik, Erin Robinson, Shelley Knuth, David Bassendine.

Collaboration will be key to the success of the Clearinghouse efforts. Those listed above as contributors will act as the primary Clearinghouse project team whose areas of collaboration should include, but are not limited to the following:
- Decisions about how to apply the Schema.org metadata scheme to the Clearinghouse learning resources with respect to:
  - Metadata template
  - Controlled vocabularies
  - Requirements & recommendations on use of the fields within the schema
  - Quality control criteria
  - Workflow recommendations
- Documentation for the Clearinghouse website and metadata form, including:
  - Glossary of terms & definitions
  - Example values
  - Help information / online guides
- User experience feedback upon Clearinghouse website content and metadata input form
- Plan for identifying sources for learning resources to add to the Clearinghouse
- Marketing and outreach for the Clearinghouse, and for contributions to it.

Communication on issues related to the Clearinghouse project will be done by means of the ESIP Data Management Training (DMT) working group under the auspices of ESIP's Data Stewardship Committee including:
- Use of the ESIP wiki, listserv, web conferencing, and Drupal content management system infrastructure
- Communication to and from ESIP's other relevant groups including the Drupal Support Working Group, and the Education Committee
- Liaisons with interested organizations by the members of the Clearinghouse project team and PIs

Sustainability

Organizing / advisory group experience:
The primary organizations collaborating on the Clearinghouse effort have been involved in producing learning resources for their constituents for a number of years. Those efforts have used a combination of authors, reviewers, editors, coordinators, and advocates, both paid and volunteer, to produce, manage and market the learning resources with limited funding. With this experience and the commitment to collaborate, representatives from each of the organizations bring a perspective that will enable them to think creatively, and make the decisions necessary to initiate and further develop the Clearinghouse so that it can continue effectively and sustainably.

Within most of the organizations involved, there are a number of staff and/or organizational resources that can bring ideas and in-kind resources to bear for the Clearinghouse. For instance, the USGS has the data management website, and the CDI itself with interest and experience in both data management training and organizational / website development. Similarly, the ESIP Federation, with its Data Stewardship Committee, Data Management Training Working Group, Drupal Support Working Group and Education Committee, is available to provide support. DataONE offers a number of successful models for both organizational and service development by means of their Community Engagement and Outreach Working Group and overall distributed leadership and governance team. For instance, this group has experience implementing online and in-person training activities, and has conducted broad community surveys in the area of data management experience and access to training. There are other organizations with which we intend to work more closely that have a great deal of experience in both data management and organizational / web services development including the National Center for Atmospheric Research (https://ncar.ucar.edu/), the University of Colorado Boulder, Research Computing Center (https://www.rc.colorado.edu/), the Research Data Alliance (https://rd-alliance.org/groups), the National Center for Ecological Analysis and Synthesis (https://www.nceas.ucsb.edu/), Software Carpentry (http://software-carpentry.org/) and its partner organization Data Carpentry (http://www.datacarpentry.org/).

Crowdsourcing / vetting process:
In early discussions about how to structure a Clearinghouse, it became immediately clear to the group that in order to make the Clearinghouse sustainable, it was going to be necessary to provide some mechanism for crowdsourcing the identification, description, and currency of the learning resources in the Clearinghouse. Fortunately, there are effective models for doing that upon which the Clearinghouse effort can build, such as that of Wikipedia, and open source software and services, such as Drupal.

During the first stage of the project, we will develop the metadata template and help resources, and make them available to those interested in contributing to the Clearinghouse. In addition, we plan to include the capability of rating the learning resources, at least by the editor and/or peer reviewers. During the CDI funded stage of the project, the processes used for editing and peer reviewing will be developed to mitigate the inherent risk involved in crowdsourcing. Experience with the process and the current hosting environment on Drupal will enable the organizing group to assess how effectively both the processes work from a sustainability point of view.

There may well be other Drupal modules available or in development that could help us achieve our sustainability goals, such as adding user-voted ratings. The feasibility of adding user-voted rating will be explored and included in a new features section of the final report.

Hosting environment:
The ESIP Federation uses the Drupal content management system extensively for all the content that it creates, including presentations and posters, and wiki pages for the many collaborative committees, clusters, working and other community groups that work within ESIP. Because of this, the Federation has a great commitment to the continuing maintenance and development of its Drupal content management system. The site is maintained by an experienced Drupal developer and managed by ESIP leadership with recommendations made by a Drupal support working group of interested and experienced Drupal developers and users within ESIP, and who network with the larger Drupal community.

Through collaboration with DataONE, the Clearinghouse will benefit from access to the broad DataONE community for outreach and promotion purposes, may leverage the experience of the DataONE Community Engagement and Outreach Working Group (co-chaired by Amber Budden and Viv Hutchison), will benefit from reciprocal links on the DataONE website and will have opportunity to participate and present at the annual DataONE Users Group Meeting which is co-located with the ESIP Summer meeting each July.

Final Report:
Given that the duration of the CDI grants are only 6 months, we anticipate that we will only be able to establish basic services for the Clearinghouse, i.e., hosting and display of the metadata for the learning resources, basic browsing and faceted searching, and an agreed upon metadata template with key help mechanisms. We already anticipate that there will be additional features and services that we will want to add to the Clearinghouse in order to make it more sustainable and useful in the future. So, for purposes of both evaluation and future planning, we will submit an internal report on the actions taken, but also include a feasibility study discussing other features and services that we anticipate adding to the Clearinghouse. Issues on sustainability that we anticipate discussing include:
- Whether and how to add to the metadata schema for purposes of establishing citations for the learning resources;
- Whether and how to use evaluation criteria from organizations offering authoritative advice on data management training core principles;
- How best to incorporate user comments on resources in the drupal environment, e.g., by author/ peer review rating, user voting, adding comments associated with user logins, accepting corrections to the metadata;
- How and whether to apply evaluation criteria for outdated learning resources to keep them useful;
- Whether there are ways to harvest and ingest MD automatically from known sources with APIs;
- Whether and how the Clearinghouse search interface can be loaded on partner web pages for more seamless integration, promotion and adoption.

Budget Justification

Funds allocated for salaries will cover tasks including project management, template creation and developer support. Nancy Hoebelheinrich will be the project manager, liaison with the technical developer and staff with the Clearinghouse project team. David Bassendine will be the technical developer for the Clearinghouse. Sophie Hou will be a contractor working on content creation, editing and production. Tamar Norkin, JC Nelson and John Faundeen will provide USGS support and guidance to the project. Travel expenses are included for 1 person to present at ESIP meetings and a CDI event (or other data meeting).

Timeline

	Objectives	Timeline
1	Establish the Clearinghouse as a subdomain on ESIP's Drupal content management program, including the development of text for the landing pages, and the mechanisms to support the workflow described below.	April 2016
2	Create a template, context-sensitive and online help guides to facilitate and promote consistency in the description of DMT resources. The template will associate the resources with the USGS Science Data Lifecycle model, and Schema.org metadata schema including the Learning Resource Metadata Initiative (LRMI) specification.	April 2016
3	Develop criteria and timelines for peer review and editor evaluations of the metadata descriptions, and refine workflow process for creating, peer-reviewing, editing, and publishing metadata descriptions for the learning resources.	April 2016
4	Using the template and workflow processes established, incorporate the descriptions of the DMT learning resources into the host environment and make them available for searching and browsing.	May-August 2016
5	Assess and report on the submission, evaluation and sustainable review process for the DMT resources, and of additional, desired features for the Clearinghouse.	September 2016

Budget form will be attached Excel table in pdf ([downloadable Excel table here](#))

Community For Data Integration (CDI) RFP BUDGET FORM		
Budget Category	Federal Funding	Matching Funds
	"Requested"	"Proposed"
1. PERSONNEL (SALARIES including benefits):		
Personnel :		
JC Nelson, 120hrs at $51.80/hr	$4,144	$2072
Tamar Norkin, 80hrs at $25.70/hr	$0	$2056
John Faundeen, 80hrs at $70/hr	$2,800	$2800
Contract Personnel :		
Nancy Hoebelheinrich, 200hrs at $75/hr	$15,000	$
Sophie Hou, 160hrs at $50/hr	$ 8,000	$
David Bassendine, 85hrs at $90/hr	$ 7,650	$
DataOne Person 40hrs at $90/hr	$0	$3600
Total Salaries:	$37,594	$10528
2. TRAVEL EXPENSES:		
Trip 1 (CDI Workshop or RDAP16, 4 days, 1 travelers)		
Per Diem:	$828	$828
Transportation (Airfare + Mileage/Shuttle):	$600	$600
Other expenses (e.g. registration fees):	$0	$0
Trip 2 (ESIP Summer Meeting, 4 days, 1 travelers)		
Per Diem:	$704	$
Transportation (Airfare + Mileage/Shuttle):	$600	$
Other expenses (e.g. registration fees):	$0	$830
Total Travel Expenses:	$2732	$2258
3. OTHER DIRECT COSTS: (itemize)		

Equipment (inc. software, hardware, purchases/rentals):	$	$
Publication Costs:	$	$
Office supplies:	$	$
Training:	$	$
Other expenses (specify):	$	$
Total Other Direct Costs:	$0	$0
Total Direct Costs:	$40,326	$16386
Indirect Cost (18.62%):	$ 7509	$
GRAND TOTAL:	$47,835	$16386

4B. Content for the Budget Template from the SOI

Budget Category	Federal Funding "Requested"	Matching Funds "Proposed"
1. PERSONNEL (SALARIES including benefits):		
Federal Personnel Total:	$10,000	$5,000
Contract/Collaborator Personnel Total:	$20,000	$5,000
Total Salaries:	$30,000	$10,000
2. TRAVEL EXPENSES:		
Travel (Per Diem, Airfare, Mileage/Shuttle) **x** # Trips:	$2,000	$2,000
Other Expenses (e.g. Registration Fees):	$0	$830
Total Travel Expenses:	$2,000	$2,830
3. OTHER DIRECT COSTS: (itemize)		
Equipment: hosting and hardware allocation	$4,000	$0
Publication Costs:	$0	$0

Office Supplies, Training, Other Expenses (specify):	$0	$0
Total Other Direct Costs:	$4,000	$0
Total Direct Costs:	$36,000	$12,830
Indirect Costs (18.62%):	$6,703.20	$2,388.95
GRAND TOTAL:	$42,703.20	$15,218.95

Data Management Plan

Data/Metadata

Data Inputs	
Title	Metadata about data management learning resources
Description	Metadata about learning resources will be collected and publicly shared via the Clearinghouse website
Format	Descriptive information saved to the Federation for Earth Science Information Partners' (ESIP's) Drupal environment
Source	The metadata will be collected from the organizations that created the learning resources, then entered into the system and reviewed by the project team
Access/sharing	Once metadata have been reviewed by project team members, the information will be accessible to the public
Version/Accessed date	Drupal includes a revisioning system that logs when a learning resource is created or modified. Every revision is timestamped and tied to a user account, and differences between revisions can be reviewed by editors in the admin interface.
Restrictions	Publicly available metadata will not have use restrictions
Data Volume Estimate	The data volume will be quite small. We can estimate that metadata records will be a few KB each. A very high estimate of records that we can get during the CDI-funded stage of the project would be 1000 records, so the largest data volume estimate we might expect would be a few MB.
Purpose	The metadata will become the content for the Clearinghouse
Quality Checks	Members of the project team will review the descriptive information about learning resources before it is made public in the Clearinghouse. The details of the quality control checks that will be used during the second phase of the project (when a crowdsourcing mechanism will be added) are still to be determined, but we will set up a system in which crowdsourced information is reviewed by a project team member before it becomes public. The system will automatically check for broken links.
Data Processing & Scientific Workflows	Does not apply

Backup & Storage	Backup and storage will be handled by ESIP's hosting environment, Pantheon (http://pantheon.io), which is a Drupal-specific hosting platform service in the cloud. The ESIP site is currently set up for automated nightly and weekly backups. Daily backups are retained for one week and weekly backups are retained for one month.
Metadata	Does not apply
Link or identifier	Does not apply
Data Publishing -- The answers entered into the "data/metadata" section apply to this section as well	

Digital/Electronic Products

Software Applications/Tools/Code -- The Clearinghouse will be built within an existing Drupal architecture. It will not require new software.

Publications

Website/Open File Reports/Fact Sheet/Press Releases/Blogs/Etc.	
Title	Final Report: Description of project deliverables and feasibility study for additional development
Description	For purposes of both evaluation and future planning, we will submit an internal report on the actions taken, and will include a feasibility study discussing other features and services that we anticipate adding to the Clearinghouse.
Link or identifier	One option is to make this report available through the ScienceBase project page that will describe the project and its deliverables.

John C. Nelson

Office: 2630 Fanta Reed Rd. La Crosse, WI 54603
Upper Midwest Environmental Sciences Center, USGS
(608) 781-6370 jcnelson@usgs.gov

EDUCATION:

B.S. in Geography, University of Wisconsin – La Crosse, La Crosse, WI 2000
    Concentrations: GIS, Cartographic Design, Population Dynamics

PROFESSIONAL EXPERIENCE:

Data Manager, 2014 – Present
    Upper Midwest Environmental Sciences Center, USGS
    -Data stewardship and metadata creation for Center
    -Performing Quality Control/Quality Assurance on data
    -Developing center standards for science information management
    -Create Geospatial products for inclusion in reports, manuscripts and
     presentations

Spatial Ecology Team Leader, 2009- present
    Upper Midwest Environmental Sciences Center, USGS
    -Oversee daily team activities
    -Develop draft fiscal year budgets and work plans
    -Represent team to outside entities
    -Identify potential project and partnerships

Geospatial Biologist, 2002 – 2014
    Upper Midwest Environmental Sciences Center, USGS
    -Perform spatial analyze on various landscapes to develop metrics to help
    researchers answer questions
    -Performing Quality Control/Quality Assurance on geospatial data
    -Create Geospatial products for inclusion in reports, manuscripts and
     presentations
    -Supervise GIS Interns

PUBLICATIONS:

Larson, J.H., Richardson, W.B., Evans, M.A., Schaeffer, J., Wynne, T., Bartsch, M.,
    Bartsch, L., Nelson, J.C., and Vallazza, J.M.. 2015. Measuring spatial variation in
    secondary production and food quality using a common consumer approach in
    Lake Erie. Ecological Applications. http://dx.doi.org/10.1890/15-0440.1

Fox, T.J., Nelson, J.C., and Rohweder, J.R. 2014. Decision Support System Development at the Upper Midwest Environmental Sciences Center. USGS Fact Sheet 2014-3120. http://dx.doi.org/10.3133/fs20143021.

Larson, J. H., Frost, P.C., Xenopoulos, M. A., Williams, C. J., Morales-Williams, A. M., Vallazza, J. M., Nelson, J. C. and Richardson. W.B., 2014, Relationships Between Land Cover and Dissolved Organic Matter Change Along the River to Lake Transition. Ecosystems, doi:10.1007/s10021-014-9804-2

Larson JH, Richardson WB, Vallazza JM, Nelson JC. 2013. Rivermouth Alteration of Agricultural Impacts on Consumer Tissue δ15N. PLoS ONE 8(7): e69313. doi:10.1371/journal.pone.0069313

Larson JH, Richardson WB, Knights BC, Bartsch LA, Bartsch MR, Nelson JC, et al. 2013. Fatty Acid Composition at the Base of Aquatic Food Webs Is Influenced by Habitat Type and Watershed Land Use. PLoS ONE 8(8): e70666. doi:10.1371/journal.pone.0070666

Larson, J.H., Richardson, W.B.,Vallazza, J.M., **Nelson, J.C**. An exploratory investigation of the landscape-lake interface: Land cover controls over consumer N and C isotopic composition in Lake Michigan rivermouths, Journal of Great Lakes Research, Available online October 2012.

De Jager, N.R, Rohweder, J.J., and **J.C. Nelson**. 2012. Past and predicted future changes in the land cover of the Upper Mississippi River floodplain, USA. River Research and Applications. OnLine First

TRAINING AND WORKING GROUPS:

<u>Training</u>
USGS Data Management Workshop 2011 and 2014
USGS Leadership Intensives 2009
USGS Leadership 101 2012
USGS Leadership 201 2015
USGS Records Liaison Training 2015
USGS Project Management Training 2015

<u>Working Groups</u>
USGS Data Management Working Group 2009 – present
Midwest Region Flood Science Technical Team 2009 – present
WI Enhanced Elevation Working Group 2010 – 2013 (chair)
USGS Workforce Planning Team 2013-2014
USGS Midwest Region Project Management Development Team 2014
USGS Natural Resources Conservation Technical Committee 2014 – present
NatureServe Spatial Methodology Review Team 2015 – present

Tamar Norkin

Science Data Management
Core Science Analytics, Synthesis, and Libraries (CSAS&L)
U.S. Geological Survey
Denver Federal Center, Building 810, MS 302, Denver, CO 80225
Phone: 303-202-4220
Email: tnorkin@usgs.gov

Current Responsibilities

USGS ScienceBase
- Point of contact for USGS scientists that use ScienceBase, an online data platform, to publicly release data. Provides assistance for scientists meeting new USGS open data policies, which include guidance for metadata, data preservation and other steps of the USGS Science Data Lifecycle.
- As a member of the ScienceBase Data Release Team, develops best practices, workflows and educational materials to guide users through the ScienceBase data release process.
- As a member of the ScienceBase Usability Team, helps evaluate and plan development of the ScienceBase user interface and creates instructional material to improve user experience.

USGS Science Data Management
- Supports the activities of the Science Data Management branch of the USGS Core Science Analytics, Synthesis and Libraries program, which manages the Community for Data Integration and creates and manages USGS tools such as the Online Metadata Editor, Science Data Catalog and the Data Management Web site.
- As a member of the Community for Data Integration's Data Management Working Group, coordinates a monthly webinar on data management topics.

Previous responsibilities include data management, integration and visualization for the Wyoming Landscape Conservation Initiative (WLCI), data management for the National Geological and Geophysical Data Preservation Program (NGGDPP), and data discovery and cataloging for the Southern Rockies Landscape Conservation Cooperative (SRLCC).

Education

University of Maryland
M.S. in Sustainable Development and Conservation Biology (2008)

Amherst College
B.S. in Biology (2003)

Employment History

2012 – Present U.S. Geological Survey, science data management
2009 – 2010   Arava Institute for Environmental Studies, executive assistant
2006 – 2007   University of Maryland, teaching assistant
2004 – 2005   Duke University, Ecology Department, research technician

Nancy J. Hoebelheinrich
448 East Ellsworth Court
San Mateo, CA 94401
650-340-1368 (home)
650-302-4493 (mobile)
njhoebel@gmail.com
nhoebel@kmotifs.com

_____

Summary of Qualifications:

- Information analyst with unique combination of knowledge and skills in working with a variety of communities across the academic, private sector, governmental, and non-profit domains on issues, problems and solutions related to the lifecycle of the digital resource from creation, retrieval and re-use, to intermediary and archival storage.
- Over 15 years of experience working as a liaison and team builder among data creators, data managers, software developers, and data users defining, designing, developing and documenting user needs, metadata input / management tools, and archiving requirements.
- Metadata subject expertise in digital library, geospatial data, government document, social science data sets, and preservation and archiving arenas.
- Experience working with GIS and remote sensing data and software (ArcGIS, GeoMedia Pro and Idrisi's Taiga), HTML, XML and RDF (various tools for web page creation & editing, schema development, XSLT, XHTML and OWL creation), data visualization tools (Adobe Acrobat and Illustrator), SQL, XQuery, MS Office applications (MS Word, Powerpoint, Excel, Access, Visio), audio and video production tools including Audacity and ProTools, and content management tools including Drupal, dSpace and Fedora.
- Experience working with social media technologies including Facebook, LinkedIn, SlideShare, Twitter, and Vimeo.

Employment Experience:

September 2009 – present   Principal Owner, Knowledge Motifs LLC
                                                     San Mateo, CA

Responsibilities:
As one of the principal owners of a small business, Knowledge Motifs LLC, provide quality consulting, data management and analysis services to educational, legal, governmental, and business enterprises needing assistance in creating, managing, and preserving data and metadata stores of and for their organizations. The company founders are veterans in the library and information management arena and offer strategy consulting, content analysis, research support, project management, and training services. Current and former clients include the Library of Congress, Association of Research Libraries, Metaweb/Freebase, Stanford University, ITHAKA, Truman Technologies LLC, the ESIP Federation, and the California Digital Library. See http://kmotifs.com/ and http://www.linkedin.com/pub/nancy-hoebelheinrich/3/6ba/57.

September, 2000 – 2009   Metadata Coordinator
                                                     Stanford University Libraries

Responsibilities:
- Analyze SUL's image, full text, audio and video collections, and geospatial and social science data sets to establish metadata requirements for content use by means of search and delivery, preservation, and digital asset management systems. Design, create and develop supporting

schemas, establish functional specifications, and define data structures for tools, modules and services associated with the Stanford Digital Repository and the Digital Library Stacks.

- Serve as a key resource to Digital Library Systems and Services by identifying requirements, strategies, tools, processes and procedures for metadata creation, capture, quality assurance and enhancement for all incoming Digital Production Group collections, and Digital Library Structure systems and services.
- Supervise the work of metadata support staff which may include ad hoc teams of regular staff, term staff, hourly staff or consultants. Coordinate with managers in Digital Library Systems & Services and the Technical Services Metadata Unit, and with other Library Technical Services and Collection Services staff as relevant to particular projects.
- Conduct research and trials to determine the most effective software tools related to metadata as the field changes.
- Provide metadata subject matter expertise to SULAIR, Stanford and the national and international metadata community by tracking and contributing to national and international information standards development committees, technical boards, and advisory groups.

June 1993 – August 2000   Technical Services Librarian
                               University of San Francisco Law School Library
                               San Francisco, CA

Responsibilities:
- Manage technical library operations including collection development, supervision of acquisitions, serials control, cataloging and database maintenance activities, binding and processing, and technology support for the automated information systems and services of the Law Library.

Selected Participation in Metadata Standards and Digital Resource Management Best Practice Development:

Member   ESIP Federation Data Stewardship Committee participant 2009 – present. Served as Co-Editor of a Data Management for Scientists Short Course consisting of 32 modules on topics related to data management planning from July 2011 – March 2013. See: http://commons.esipfed.org/datamanagementshortcourse Authored 3 modules; separate links can be found at my LinkedIn profile above.

Member   IEEE Learning Technology Standards Committee RAMLET Working Group to develop a conceptual model for content packaging representing METS, December 2004 – present. See: https://mentor.ieee.org/ramlet/documents

Member   Editorial Board, Metadata Encoding & Transmission Standard (METS), sponsored by Digital Library Federation (DLF), May 2001 – present; Administrative Co-Chair April 2006 - 2012.

Member   PREMIS (PREservation Metadata Implementation Strategies) Working Group, sponsored by OCLC, and the Research Libraries Group (RLG), Implementation Subgroup. April 2003 – 2005.

Education:
- Foothill College, Los Altos Hills, CA, **Certificate of Achievement for Geographic Information Systems Analyst**.
- University of Washington, Seattle WA, **Master of Library & Information Science**.
- Stanford University Interdisciplinary Program in Social Theory, **BA** *with distinction*.

NCAR Library

NCAR

National Center for Atmospheric Research
NCAR Library
1850 Table Mesa Drive
Boulder, CO 80305
Web: http://www.library.ucar.edu

January 20, 2015

Nancy Hoebelheinrich
Knowledge Motifs
San Mateo, CA 94401

Dear Ms. Hoebelheinrich:

I write to express my interest in collaborating with your project team in the development of this Data Management Training Clearinghouse. Data management training resources are emerging from many scientific communities in response to a common set of goals and constraints, namely, that digital data and computation provide new means of conducting analytical research, and that dealing with such data is a complex and resource intensive task.

In my role as the Research Data Services Specialist at the National Center for Atmospheric Research (NCAR)/University Corporation for Atmospheric Research (UCAR), located in Boulder, CO, I coordinate the NCAR Library's data-related initiatives. From 2012-2014 I coordinated the NCAR Library's contributions to the Data Curation Education in Research Centers program (DCERC), which was a partnership between NCAR, the University of Illinois at Urbana-Champaign, and the University of Tennessee, Knoxville. DCERC supported nine Master's and three Ph.D. students to conduct data curation internships at NCAR. I also contributed to the development of the Earth Science Information Partners (ESIP) Data Management Short Course, authoring eight training modules on multiple data management topics.

For this project, I will provide guidance of the types of data management resources that this Clearinghouse might collect, based on my experiences with DCERC, ESIP, and other projects. I look forward to this Clearinghouse filling a significant gap with a much-needed service.

Sincerely,

Matthew S. Mayernik, Ph.D.
Project Scientist & Research Data Services Specialist
NCAR Library / UCAR Integrated Information Services
National Center for Atmospheric Research (NCAR)
University Corporation for Atmospheric Research (UCAR)
Boulder, CO
mayernik@ucar.edu

22 Jan 2016

Nancy Hoebelheinrich
Knowledge Motifs LLC
San Mateo, CA 94401

Dear Nancy,

I am pleased to write a DataONE letter of support for the proposed Data Management Training Clearinghouse. This collaborative project will provide the Earth science community with needed services cataloging and easily discoverable training resources for data management and education. Coordination of data management education efforts will remove the barriers associated with data archiving and preservation and will facilitate future research that is based on openly accessible data.

DataONE will contribute to the Clearinghouse by providing the broad DataONE education resources, by promoting the Clearinghouse within the DataONE community, and by contributing to research and development products undertaken by the team. Amber Budden, Director of Community Engagement and Outreach and co-chair of the DataONE Community Engagement and Outreach Working Group, will contribute to the Clearinghouse project and will oversee communication between the Clearinghouse and DataONE.

We in DataONE look forward to working with you in this collaboration and eagerly anticipate its success.

Sincerely,

William Michener
Professor and Director of e-Science, University Libraries, University of New Mexico
Principal Investigator, DataONE (Observation Network for Earth)

January 22, 2016

Nancy J. Hoebelheinrich
Knowledge Motifs
San Mateo, CA 94401

Dear Ms. Hoebelheinrich:

I am pleased to submit this letter of support from the Foundation for Earth Science (FES) on behalf of the Federation of Earth Science Information Partners (ESIP) for your proposal to the USGS CDI, "Data Management Training Clearinghouse".

As an organization, the ESIP Federation has an 18-year track record of enabling Earth science interoperability by seeding innovation, developing best practices and advancing technologies across broad-based, distributed communities of science, data and information technology practitioners. ESIP optimizes collaboration through backbone infrastructure including in-person meetings and virtually through collaboration spaces on the Web to support community-defined topics.

The ESIP Federation's status as a leading collaboration network has made it the go-to place to forge consensus on emerging data-related topics, especially around data management training. ESIP's membership includes federal data centers, government research laboratories, research universities, education resource providers, technology developers, and various nonprofit and commercial enterprises. Beyond formal membership, ESIP is connected to other networked organizations such as the Research Data Alliance (RDA), DataOne, World Data Systems (WDS), the Group on Earth Observations (GEO), and a long-standing partnership with NASA and NOAA initiative through common members and shared activities.

We recognize that the Data Management Training Clearinghouse is a significant contribution to the geoscience community and will require collaboration from a variety of stakeholders. We look forward to partnering with you to leverage our collective communities and resources through developing this shared resource.

We believe that by virtue of working together, ESIP and USGS CDI will experience the network effect, enabling a scalable training model across domain-specific communities and benefiting science in ways we can only imagine by making data matter.

Sincerely,

Erin Robinson
Executive Director
Foundation for Earth Science/ ESIP Federation
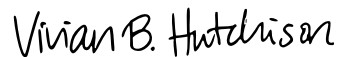
January 21, 2016

Dear Reviewers,

The Community for Data Integration (CDI) Data Management Working Group is pleased to submit this letter of support for the CDI proposal entitled "Data Management Training Clearinghouse" (submitted by J.C. Nelson and colleagues).

The main objective of the Data Management Working Group is to develop mechanisms for incorporating data management into USGS science and develop ways to educate scientists of its value. The group seeks to elevate the practice of data management such that it is seen as a critical partner in the pursuit of science in USGS.

The Data Management Training Clearinghouse proposal represents a strong idea for a welcome tool for gathering information in one place such that it is easier to locate, access, and evaluate for specific training needs. The Working Group looks forward to seeing the benefits afforded by such a resource and are pleased to offer a letter of support towards that end.

Sincerely,

Vivian B. Hutchison

Viv Hutchison
Co-lead, Community for Data Integration Data Management Working Group